Background on Deep Learning

Nate Merrill

University of Delaware

Popular Deep Learning Tasks

- Classification
- Classifcation + Bounding Box Detection
- Muti-object detection
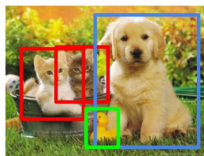- Pixel-wise semantic segmentation

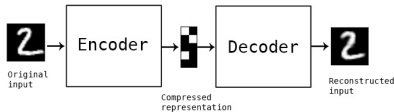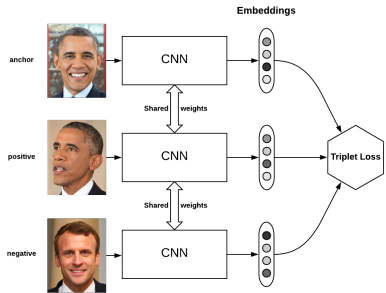| Classification	Classification + Localization	Object Detection	Instance Segmentation
CAT	CAT	CAT, DOG, DUCK	CAT, DOG, DUCK
Single object		Multiple objects	

More Popular Deep Learning Tasks

- Embeddings (dimension reduction)
  - Autoencoder
  - Triplet Embedding

Autoencoder

Triplet embedding
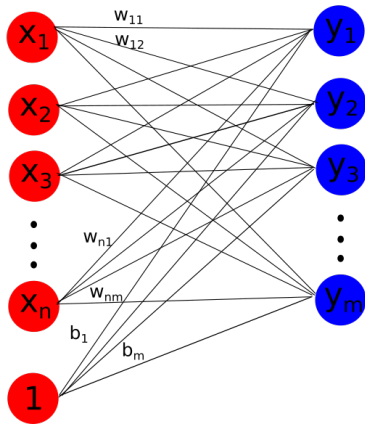
Fully Connected Layer

- Vector input and output
- Learned weights associated with each connection
- Can be written as a linear operation

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n & 1 \end{bmatrix}^\mathsf{T}$$

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_m \end{bmatrix}^\mathsf{T}$$

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1n} & b_1 \\ w_{21} & w_{22} & \ldots & w_{2n} & b_2 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ w_{m1} & w_{m2} & \ldots & w_{mn} & b_m \end{bmatrix}$$
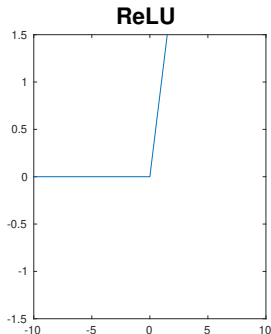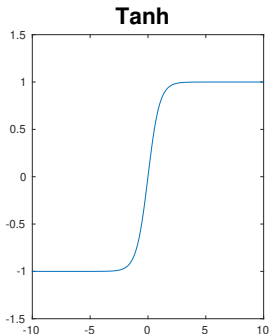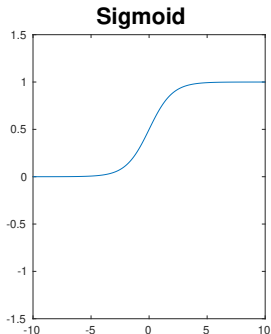
$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Nonlinear Activation

- Fully connected layer cannot model nonlinear functions
- Nonlinear activations are used to provide nonlinearity

Examples:

Popular activation functions

Convolution Layer

- Matrix/Tensor input and output
- Useful for image input

$$
\begin{array}{|c|c|c|c|c|c|c|}
\hline
X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} & X_{17} \\
\hline
X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} & X_{27} \\
\hline
X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} & X_{37} \\
\hline
X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} & X_{47} \\
\hline
X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} & X_{57} \\
\hline
X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} & X_{67} \\
\hline
X_{71} & X_{72} & X_{73} & X_{74} & X_{75} & X_{76} & X_{77} \\
\hline
\end{array}
\ast
\begin{array}{|c|c|c|}
\hline
K_{11} & K_{12} & K_{13} \\
\hline
K_{21} & K_{22} & K_{23} \\
\hline
K_{31} & K_{32} & K_{33} \\
\hline
\end{array}
=
\begin{array}{|c|c|c|c|c|c|}
\hline
Y_{11} & Y_{12} & Y_{13} & Y_{14} & Y_{15} & Y_{16} \\
\hline
Y_{21} & Y_{22} & Y_{23} & Y_{24} & Y_{25} & Y_{26} \\
\hline
Y_{31} & Y_{32} & Y_{33} & Y_{34} & Y_{35} & Y_{36} \\
\hline
Y_{41} & Y_{42} & Y_{43} & Y_{44} & Y_{45} & Y_{46} \\
\hline
Y_{51} & Y_{52} & Y_{53} & Y_{54} & Y_{55} & Y_{56} \\
\hline
Y_{61} & Y_{62} & Y_{63} & Y_{64} & Y_{65} & Y_{66} \\
\hline
\end{array}
$$

$$
\begin{aligned}
Y_{ij} = {} & K_{33}X_{i,j} + K_{32}X_{i,j+1} + K_{31}X_{i,j+2} + K_{23}X_{i+1,j} \\
& + K_{22}X_{i+1,j+1} + K_{21}X_{i+1,j+2} + K_{13}X_{i+2,j} \\
& + K_{12}X_{i+2,j+1} + K_{11}X_{i+2,j+2}
\end{aligned}
$$

Max Pooling Layer

- Useful for viewpoint invariance
- Similar operation to convolution